General introduction

In this chapter we will introduce the key players of our research - enzymes called glucansucrases and fructansucrases - and provide the context for the investigations described in this thesis.

Plants and bacteria are able to synthesize complex, extracellular carbohydrate polymers from a simple sugar, sucrose. These extracellular polysaccharides (EPS) are essential for the organisms to thrive in their natural environment by functioning as an energy reserve, as well as enabling them to colonize host organisms and protect themselves against all kinds of environmental stress.

After describing the different types of EPS, as well as their use in food, health and other industrial applications, we turn to the molecular ‘factories’ that produce EPS, glucansucrase and fructansucrase enzymes, especially those found in *Lactobacillus* species. Current knowledge about these enzymes will be summarized, followed by the scope of this thesis - to provide some pieces in the puzzle that is called the structure-function relationship.

1. CARBOHYDRATES

1.1. Biological function of carbohydrates

The best known carbohydrate is probably sucrose (saccharose, Figure 1) - commonly known as ‘sugar’. Carbohydrates are a class of biomolecules represented by the chemical formula \( \text{C}_n(\text{H}_2\text{O})_n \). Varying greatly in complexity and size, carbohydrates represent one of the four major classes of biomolecules together with lipids, proteins and nucleic acids.

\[ \text{Figure 1. Structure of sucrose, consisting of a glucose moiety (left) and a fructose moiety (right); the full name is } \beta-\text{D-fructofuranosyl } \alpha-\text{D-glucopyranoside.} \]

Given their variety and widespread occurrence in nature, it is not surprising that carbohydrates have a great variety of functions essential to living organisms (Berg et al., 2007, p303). First, they provide an important source of energy. For example, carbohydrates found in our food are processed by our digestive system to smaller carbohydrates such as glucose, and then taken up by our body to provide energy for our cells, tissues and organs. Any carbohydrate that is not needed immediately is stored in a special form for later use. Like humans, also plants and bacteria use carbohydrates as an energy source; plant cells for example store carbohydrate polymers, a reserve which is addressed after winter dormancy when leaves need to be (re)grown (Banguela & Hernández, 2006). A second function of carbohydrates is in communication between cells; hence their role in the immune system, and in infection and inflammation processes. For example, ovulated egg cells in mammals are surrounded by a coat of glycoproteins carrying specific carbohydrate ‘antennas’; when a sperm cell recognizes these carbohydrates, it releases enzymes that dissolve the coat and allow sperm entry and fertilization. Another example is the decoration of protein antigens in human serum with varying carbohydrates, which are recognized by antibodies, thus defining the ABO-blood group system. The number of carbohydrate antenna types is vast, and understanding the interactions they make is of major importance for diagnosing diseases and for the development of drugs. A third function of carbohydrates becomes obvious when we look at the living nature around us: they provide structural integrity and protection. The cell walls of plants, animals, bacteria and archaea are largely made up of a network of carbohydrate polymers; this network gives the organisms structure and strength, while maintaining flexibility.
Some organisms produce a layer of carbohydrates outside the cell wall, called the exopolysaccharide (EPS) layer. EPS-producing strains have been identified in a large variety of biological systems (fermenting cabbages, fruits, vegetables, sourdoughs, beverages, dairy, cereals, dental plaque, intestines) and in environmental spills of sugar factories (Leemhuis et al., 2013a). In this thesis we will focus our attention on specialized enzymes that synthesize EPS carbohydrates in certain lactic acid bacteria (LAB). The EPS layer may serve as an energy reserve when other sources of energy are scarce. Importantly, it also protects the bacteria against hostile organisms or toxic compounds, and against environmental stress such as drought or low temperatures (Flemming & Wingender, 2010). This protective function is further enhanced in so-called biofilms, dense layers of microorganisms surrounded by EPS. Due to the slimy and sticky properties of the EPS, which may represent up to 90% of the total mass, these biofilms easily adhere to other surfaces. In this way, the EPS helps the bacteria to thrive attached to or even inside host organisms. A well-known example of a biofilm is found in the human oral cavity, where \textit{Streptococcus mutans} produces an insoluble type of EPS, sticking to tooth surface; because other harmful bacterial species easily attach to this biofilm, this eventually may result in dental plaque and tooth decay (Balakrishnan \textit{et al.}, 2000; Ito \textit{et al.}, 2011; Rozen \textit{et al.}, 2001). A more beneficial type of microbial community is found in the human intestinal system, which is colonized by different strains of bacteria, which help to digest nutrients that cannot be processed by the human gastric enzymes. Importantly, LAB have beneficial effects on the intestinal microbial ecosystem by suppressing growth of pathogens. Examples of such human symbionts are \textit{Lactobacillus reuteri} and \textit{Lactobacillus gasseri}, the predominant autochthonous (indigenous) species in the human intestinal tract (Reuter \textit{et al.}, 2001), and \textit{L. johnsonii} (Anwar, 2010a). The mechanisms by which LAB suppress growth of pathogenic species have not been clearly established; however the EPS produced by LAB is believed to play an important role for their functioning in the host (Kleerebezem \textit{et al.}, 2010). Therefore, knowledge about the processes and mechanisms of EPS-synthesis by these organisms may be used advantageously in the development of food- and health-related applications.

1.2. Applications of EPS

During the 19th century, it was discovered that bacteria produce sugar-like compounds responsible for the thick texture of sugar cane and beet syrups (Crescenzi, 1953; Pasteur, 1861). The first species found to do so was the lactic acid bacterium \textit{Leuconostoc mesenteroides}, which produces large carbohydrate polymers (polysaccharides) that were termed dextrans. Later, many more bacterial species were found to produce many different types of (exo)polysaccharides, roughly divided in $\alpha$-glucans and $\beta$-fructans.
The best known and commercially most widely applied EPS are dextrans, belonging to the α-glucan type. In the food industry, dextrans are used because of their gelling, viscosifying, and emulsifying properties in baked products, dairy products and beverages (Leemhuis et al., 2013a). For example, *in situ* production of dextrans in sourdough has been applied to improve the texture, volume and crumb softness of breads (Palomba et al., 2012). In addition, dextran-related gluco-oligosaccharides (GOS) are commercially used as low-calorie sweeteners (Carlson et al., 2009; Carlson et al., 2011; Grysman et al., 2008). Dextrans also have numerous health-related applications, such as their use as a blood plasma expander or antithrombotic agent, and their use to prevent surface colonization by pathogenic bacteria (Wang et al., 2010). Furthermore, they can be used as bioactive agents (e.g. as immunomodulators, anti-ulcer agents or cholesterol-lowering agents), metal-complexing agents, emulsifiers, detergents, bleaching agents, anti-corrosive agents, strengthening additives, and in waste-water purification and cosmetic applications (Leemhuis et al., 2013a; Van Geel-Schutten et al., 1999).

Like dextrans and related α-glucans, also EPS of the β-fructan type find many food- and health related applications (Kaur & Gupta, 2002). For example, β-fructans and related fructo-oligosaccharides (FOS) are used in food products such as desserts, fermented dairy, and infant formulae; they can replace fat, be used as natural sweeteners (e.g. the β-fructan 1-kestose), or improve the texture and stability. In non-food applications, β-fructans are used as a biological glue (http://www.polysaccharides.us/levanadhesive_summary.php), as biodegradable surface-active agents (Stevens et al., 2001), or as protective agents for artificially generated human skin (Kim et al., 2005).

Due to the unique properties of EPS-related poly- and oligosaccharides from LAB, they have gained substantial interest in recent years. Being difficult to produce in sufficient quantities and reasonable prices from biomass (Rastall & Hotchkiss, 2003), methods for production of EPS-like oligo- and polysaccharides by enzymatic synthesis are investigated (Maiorano et al., 2008). By engineering either the enzymes or the substrates, even tailor-made oligosaccharides with altered structures and with desired properties can be synthesized (Hellmuth et al., 2008; Homann & Seibel, 2009b; Kang et al., 2011; Kelly et al., 2009; Kralj et al., 2008a). Such products can be applied in e.g. (medical) diagnostics, screening of vaccines, and cell-surface labeling (Homann & Seibel, 2009a).

### 1.2.1. Prebiotics and probiotics

A relatively new and rapidly growing area of research is the prebiotic effect that EPS α-dextrans and β-fructans may have. The term prebiotic applies to compounds that have a health-promoting effect by stimulating the growth of beneficial endogenous microbiota
in the (human) gastrointestinal tract (German et al., 1999; Welman & Maddox, 2003). Prebiotic compounds are not degraded by the digestive enzymes in the upper gastrointestinal tract, and thus have a low caloric value and dietary fiber-like properties. Instead, \( \alpha \)-dextrins and \( \beta \)-fructans are metabolized by beneficial resident colon bacteria, resulting in the proliferation of healthy gut bacteria while pathogenic bacteria are competed out. This in turn leads to positive health effects such as lower blood cholesterol and triacylglycerol lipid levels, B-vitamin production, increased resorption of calcium and magnesium ions, and prevention of colon cancer (Gibson et al., 1995). Bacterial strains that have health-promoting properties for the human gut flora are termed probiotic. With more and more known EPS-producing strains being identified, it is not surprising that probiotics and prebiotics have become of high interest to the food industry. The marketing of dairy products such as Yakult, Actimel and Activia, containing probiotic \textit{Lactobacillus} or \textit{Bifidobacterium} cultures, exemplifies this interest. With a proven stimulating effect on intestinal \textit{Bifidobacterium} species in humans and animals, all \( \beta \)-fructans are considered prebiotic (Banguela & Hernández, 2006). Well known examples are the natural inulin-type FOS found in vegetables such as onion, garlic, Jerusalem artichokes, asparagus, tomatoes etc.; the most potent prebiotics being the inulin-type FOSs 1-kestose, 1-nystose and 1-fructofuranosyl-nystose (Moore et al., 2003). In case of \( \alpha \)-dextrins and related carbohydrates, the prebiotic effect is already proven for some, while for many others it is still under investigation. Methods to synthesize naturally occurring and modified \( \alpha \)-dextrins and \( \beta \)-fructans are being developed with the goal to obtain products with desired (\textit{e.g.} prebiotic) properties (Homann & Seibel, 2009b; Kralj et al., 2008a; Maiorano et al., 2008).

### 1.3. Structures and properties of EPS

The classification of exopolysaccharides produced by certain lactic acid bacteria is based on the sugar residue units of which they are composed. Heteropolysaccharides consist of different types of units (mainly glucose, galactose, rhamnose or fructose), and may be decorated with acetate, phosphate or glycerolphosphate (De Vuyst & Degeest, 1999). Homopolysaccharides consist of a single type of unit, either glucose (\( \alpha \)-glucans) or fructose (in \( \beta \)-fructans). Further classification of hetero- and homopolysaccharides is based on the way the individual units are linked via glycosidic bonds.

#### 1.3.1. \( \alpha \)-Glucans

The building block of \( \alpha \)-glucans is the monosaccharide \( \alpha \)-D-glucose (Figure 2a), with five hydroxyl groups (-OH) available for glycosidic linkage. One unit may link in four different
ways to the next with an $\alpha(1\rightarrow 2)$, $\alpha(1\rightarrow 3)$, $\alpha(1\rightarrow 4)$ or $\alpha(1\rightarrow 6)$ glycosidic linkage; the consecutive formation of glycosidic bonds (elongation) results in a polymer. In addition, glucose units in such a polymer may have a secondary glycosidic linkage, resulting in a branched chain with $\alpha(1\rightarrow 2,6)$, $\alpha(1\rightarrow 3,6)$ or $\alpha(1\rightarrow 4,6)$ linkages. The $\alpha$-glucans have been classified according to the dominant glycosidic linkage type in the main chain (Figure 2b-e). Given the many different linkage possibilities (linkage type, number of units, degree and type of branching, length and spatial arrangement of branches), it is easy to imagine that the structural variation in $\alpha$-glucan type EPS is enormous. The terminal sugar unit of $\alpha$-glucan chains is a fructosyl unit, derived from sucrose.

Conceivably, the structural variation in $\alpha$-glucans leads to a huge variation in physicochemical properties, such as solubility, viscosity and stickiness. For example, mutans are relatively insoluble, while dextrans are much more soluble. In general, $\alpha$-glucans are very flexible structures; they may reach molecular weights of up to 40 MDa, corresponding to a degree of polymerization (DP) of $\sim 250,000$. Techniques such as high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), high-resolution $^1$H and $^{13}$C NMR, and high-resolution mass spectrometry have made it possible to analyze the molecular structure of $\alpha$-glucans in more detail. This
has already been applied to the α-glucans synthesized by the glucansucrases GTF180 and GTFA from certain *Lactobacillus reuteri* strains (Van Leeuwen et al., 2008a,b,c). The composite structural models derived from such studies show that the α-glucan produced by a specific glucansucrase enzyme has a very complex but relatively well-defined and specific structure. In addition to analysis of the final products (α-glucan), initial and intermediate reaction products (oligosaccharides) of the glucansucrase reaction are also being analyzed (Côté et al., 2008, Dobruchowska et al., 2013); in this way one may follow reaction progress to gain insight in the reaction specificity of different glucansucrases, especially regarding the glycosidic linkage type specificity. A similar approach has been applied very recently to an α(1→2) branching glucansucrase (Brison et al., 2013), revealing the branching pattern of its short oligosaccharide products.

### 1.3.2. β-fructans

The repeating unit found in β-fructans is β-D-fructose, an isomer of α-D-glucose with a five-membered ring (Figure 3a). Like α-D-glucose, β-D-fructose has five hydroxyl groups, but only three of these are found to form glycosidic linkages to neighboring units. The 2-OH forms either β(2→1) glycosidic linkages in inulin, or β(2→6) glycosidic linkages in levan; due to the mechanism of β-fructan synthesis (by fructansucrases), the terminal unit of inulins and levans is an α-D-glucose, derived from sucrose.

![Figure 3](image)

**Figure 3.** Structures of linear β-fructans. (a) β-D-fructose, the building block of β-fructans; (b) inulin; (c) levan. In (b) and (c), the carbon atoms making glycosidic bonds between the fructose (and terminal glucose) units are numbered in red.

Like α-glucans, β-fructans display a large structural variation. While β-fructans produced by bacteria are largely linear and may reach DPs of 1,000,000, plants and fungal species synthesize generally shorter (DP <200) and more diverse types of β-fructan, for example
with \(\beta(2\rightarrow1)\) and \(\beta(2\rightarrow6)\) in the same polymer (mixed levan), or with fructose chains attached to both the 1- and the 6-\(\text{OH}\) group of the glucose unit (neoseries levan and inulin) (Banguela & Hernández, 2006; Van Hijum et al., 2006). It has been reported that \(\beta\)-fructans are able to form helical structures. In the case of inulin-type FOS, such helices may be left- or right-handed, and rather flexible (French, 1988; French, 1989; Marchessault et al., 1980; Valluru & Van den Ende, 2008; Vereyken et al., 2003). Levan-type FOS are more rigid, preferring left-handed helix formation. Additional structural variation arises from the fact that \(\beta\)-fructans may be branched (e.g. levans with \(\beta(2\rightarrow1)\) linked branches).

Because of their structural variation, \(\beta\)-fructans have different physicochemical properties, which are of importance for their applicability. For example, it has been suggested that the high solubility of certain \(\beta\)-fructans in water must be related to a high degree of flexibility, and that extensive branching may reduce flexibility and solubility (Pollock & Cairns, 1991).

2. GLUCANSUCRASES

2.1. Identification of glucansucrases and glucansucrase-producing strains

Glucansucrases (GSs, also called glucosyltransferases, GTFs) are extracellular enzymes that catalyze the synthesis of \(\alpha\)-glucan exopolysaccharides from sucrose (Leemhuis et al., 2013a). They are often named according to the type of \(\alpha\)-glucan they synthesize, e.g. dextran-sucrase, mutan-sucrase, reuteran-sucrase, alternan-sucrase. So far, GS have basically been identified in LAB genera *Lactobacillus*, *Leuconostoc*, *Streptococcus*, *Oenococcus* and *Weissella*. Also due to the interest for food-related applications, the number of known GS-producing bacteria is rapidly increasing. Identification of such strains is initially based on the appearance of slimy/ropy colonies or viscous cultures, which indicates that sucrases are synthesizing EPS. By replacing sucrose with raffinose, a separation can be made between strains that have FSs (which are able to use raffinose as a substrate) and those having GSs (unable to do so). Hundreds of glucan- and fructan-synthesizing bacterial strains have been identified in this way; however the number of confirmed and characterized GSs (and FSs) has lagged behind. Over 200 gene sequences have been classified as GS in the CAZY database to date (http://www.cazy.org; Cantarel et al., 2009), but so far only about 57 have indeed been confirmed to have GS activity. Most GSs have a molecular mass between 120-200 kDa, although some are even larger; an example of the latter is DSR-E from *Leuconostoc mesenteroides* NRRL B-1299 with a molecular mass of 313 kDa.
2.2. Classification and primary structure of glucansucrases

GSs are classified as glycoside hydrolase family 70 (GH70) in the CAZy system, a database that is based on amino acid sequence similarities (Cantarel et al., 2009). Within a GH family, the catalytic machinery and mechanism are conserved. The GH70 enzymes are mechanistically, structurally and evolutionarily closely related to enzymes belonging to GH13 and GH77. Together, these three families constitute the GH-H clan (Henrissat & Davies 1997; MacGregor et al., 2001; Stam et al., 2006). The common feature of GH-H enzymes is that they can cleave the α-glycosidic linkage between a glucose moiety and another sugar moiety (glucose, fructose etc.); the catalytic residues are located in a (β/α)$_8$ barrel domain. A noteworthy member of GH13 is amylosucrase from Neisseria polysaccharea, which synthesizes α(1→4) linked glucose polymers from sucrose, and thus can be regarded as a GH13 glucansucrase (Skov et al., 2000). The enzymes within the GH-H clan share four signature sequence motifs (I-IV; Figure 4); interestingly, the order of the motifs is different in GH70 enzymes due to a circular permutation of structural elements in the (β/α)$_8$ barrel.

Sequence alignment and phylogenetic analysis of all enzymatically characterized GH70 enzymes (Leemhuis et al., 2013a) shows that GSs with distinct product specificities do not form clear clusters, with the exception of a small group of enzymes sharing ~50% sequence identity with GSs, but having a substrate and product specificity that is clearly different from GSs (Kralj et al., 2011; Leemhuis et al., 2013b). This suggests that even small differences in amino acid sequence are sufficient to provide different specificities; this opens up possibilities to engineer GH70 enzymes in order to obtain desired substrate and product specificities. Indeed, Kralj et al. (2005) mutated the L. reuteri GTFA from a reuteran- to a dextransucrase by just three amino acid substitutions.

The enzymes classified within GH70 share an amino acid sequence identity of 30-80%. Based on sequence comparison with GH13 α-amylases, and before structural information was available, it was proposed that GH70 GSs, in addition to the catalytic domain, have N- and C-terminal domains of variable length, displaying a 'linear' domain organization (Figure 4a). However, the first determined structure of a (N-terminally truncated) GS (Vujičić-Žagar et al., 2010) unequivocally showed that additional domains and even the catalytic domain are constituted from two polypeptide segments, resulting in a 'U-shape' arrangement (Figure 4b, c). Only one of the additional domains (C) consists of a single contiguous polypeptide segment. Thus, the N-terminal polypeptide segment preceding the catalytic domain contributes to domains V, IV and B (Figure 4d); moreover it forms the large N-terminal domain that is lacking in the truncated structures. The C-terminal polypeptide segment following the catalytic domain contributes to domains B, IV and (although not in all GSs) to domain V.
The N- and C-terminal segments ‘outside’ the core domains have been designated ‘glucan binding domains’ (GBD), and their importance has been investigated in many studies including truncation and deletion mutants (Abo et al., 1991; Brison et al., 2012; Kralj et al., 2004a,b; Kingston et al., 2002; Lis et al., 1995; Monchois et al., 1999a,b,c). In Lactobacillus species, the N-terminal segment contains 200-700 residues. The N-terminal segments of most Streptococcus and Leuconostoc GSs are relatively short (~150 residues); exceptions are the large middle domain (839 residues) in DSR-E from L. mesenteroides NRRL B-1299 (Bozonnet et al., 2002), and the absence of an N-terminal domain in L. mesenteroides DsrA (Janeček et al., 2000). The C-terminal segment of most glucansucrases from Streptococcus and Leuconostoc species spans about 500 residues, but in Lactobacillus species it is usually shorter (e.g. 264 residues in GTF180 and GTFA).

2.3. The 3D structure of glucansucrases

Only recently the first 3D structures of GH70 GSs have become available (Table 1, Figure 5). In all cases, the crystals were obtained with truncated forms of the enzymes, containing domains A, B, C, IV (and V); crystal structures of complete GS enzymes have not yet been reported to date.

These 3D structures revealed several surprising structural features. First, the catalytic core of the GSs consists of three core domains, which resemble the A, B and C domains also found in glycoside hydrolase family 13 (GH13) enzymes (Figure 4d). Secondly, attached to the core domains are two extra domains, called IV and V. Third, and most surprisingly, instead of the expected linear domain arrangement, a ‘U-shaped’ arrange-
Table 1. Known 3D structures of GH70 glucansucrases. Scr = sucrose, mal = maltose, acb = acarbose; * = different crystal form.

<table>
<thead>
<tr>
<th>Source and linkage specificity</th>
<th>Name</th>
<th>Construct (nr. of amino acids in full length protein)</th>
<th>PDB ID</th>
<th>Resolution (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lactobacillus reuteri 180</td>
<td>GTF180-ΔN</td>
<td>741-1772 (1772)</td>
<td>3KLK</td>
<td>1.65</td>
<td>Vujičić-Žagar et al., 2010</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3HZ3</td>
<td>2.22 (scr)</td>
<td>Chapter 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3KLL</td>
<td>2.00 (mal)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4AYG</td>
<td>2.00*</td>
<td></td>
</tr>
<tr>
<td>Lactobacillus reuteri 121</td>
<td>GTFA-ΔN</td>
<td>741-1781 (1781)</td>
<td>4AMC</td>
<td>3.60</td>
<td>Pijning et al., 2012;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chapter 4</td>
</tr>
<tr>
<td>Streptococcus mutans</td>
<td>GTF-SI</td>
<td>244-1087 (1455)</td>
<td>3AIE</td>
<td>2.10</td>
<td>Ito et al., 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3AIB</td>
<td>3.09 (mal)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3AIC</td>
<td>3.11 (acb)</td>
<td></td>
</tr>
<tr>
<td>Leuconostoc mesenteroides</td>
<td>DSR-E ΔN123-GBD-CD2</td>
<td>1782-2825 (1781-2834)</td>
<td>3TTQ</td>
<td>1.90</td>
<td>Brison et al., 2012</td>
</tr>
<tr>
<td>NRRL B-1299</td>
<td></td>
<td></td>
<td>3TTO</td>
<td>3.30</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5. Published 3D structures of glucansucrases. For each glucan-sucrase, the cartoon representation as well as a schematic domain organization is given. Domains are colored as follows: domain V, red; domain IV, yellow; domain B, green; domain A (catalytic domain), cyan; domain C, violet.

Four of the five domains (Figure 4c, d) were present with four of the five domains being constituted by N- and C-terminal polypeptide segments. From the N- to C-terminus the polypeptide chain successively contributes to domains V, IV, B, A, C, A, B, IV and V. Only domain C consists of one contiguous polypeptide. The 3D structures of the domains found in GSs will be described individually hereafter.
2.4. The 3D structures of individual domains

Domain A comprises a (β/α)₈ barrel and harbors the catalytic site. The three proposed catalytic residues (the nucleophilic aspartate, the acid/base glutamate and the transition state stabilizing aspartate) lie at the bottom of a deep pocket in this domain; the crystal structure of GTF180-ΔN mutant D1025N with bound sucrose (Vujičić-Žagar et al., 2010) confirmed that the three residues constitute the catalytic site of GSs as predicted (MacGregor et al., 1996). They are in loops following β-strands β4, β5 and β7, in homology regions II, III and IV, respectively, in the N-terminal part of domain A (Figure 6). The fourth homology region I is located in the C-terminal part of domain A, confirming the predicted circular permutation of the (β/α)₈ barrel (MacGregor et al., 1996). A model of how such a permutation may have evolved has been proposed (Vujičić-Žagar et al., 2010). The active site pocket and groove are surrounded by several loops from domains A and B. Differences in these loops between the known GS structures are described in Paragraph 2.8.

Figure 6. Sequence alignment of conserved sequence motifs I-IV of GH70 enzymes. NU = nucleophile, A/B = general acid/base, TS = transition state stabilizer. Boxes ‘-1’, ‘+1’ and ‘+2’ designate residues interacting with the substrate in subsite -1 of the sucrose complex of GTF180-ΔN, or with maltose at subsites +1 and +2 of the maltose complexes of GTF180-ΔN and GTF-SI. The blue box in motif IV indicates the tripeptide targeted in several mutation studies. LrGTF180, Lactobacillus reuteri 180 GTF180; LrGTF, Lactobacillus reuteri 121 GTF; LrGTFO, Lactobacillus reuteri ATC 55730 GTF; LmDSRS, Leuconostoc mesenteroides NRRL B-512F dextrantransucrase; LmDSRBCB4, Leuconostoc mesenteroides B-1299 DSRB CB4 dextrantransucrase; LmASR, Leuconostoc mesenteroides B-1355 alternansucrase; SmGTF-SI, Streptococcus mutans glucosyltransferase GTF-SI; SoGTFR, Streptococcus oralis ATC10557 glucosyltransferase R; LmDSREG2, Leuconostoc mesenteroides NRRL B-1299 DSR-E (glucan binding domain and CD2); LrGTFB, Lactobacillus reuteri 121 4,6-α-glucanotransferase. The alignment is prepared with ESPript (Gouet et al., 1999).

Domain B forms a highly twisted antiparallel five- or six-stranded β-sheet. Located next to the catalytic domain, this domain is also essential for GS function. First, some of the loops in this domain contribute to shaping the groove near the catalytic site. Second,
residues at the interface of domain B with domain A form a calcium binding site at about 10 Å from the nucleophilic aspartate. The presence of Ca\(^{2+}\) has been shown to be essential for the activity of GTFA\(-\Delta N\) (Kralj et al., 2004b) and GTF180\(-\Delta N\) (Vujičić-Žagar et al., 2010).

**Domain C**, also located next to the catalytic domain but ‘opposite’ to domain B, has an eight-stranded β-sheet with a Greek key motif. Although it is conserved in all GS structures as well as in most GH13 enzymes, a function has not yet been attributed to this domain.

The structure of **domain IV** revealed a novel fold with no similarity to any other known protein structures. Domain IV connects domains B and V, but whether it has any other function is unknown; Ito et al. (2011) proposed that domain IV may provide a 'hinge' that would bring domain V with bound glucans near to or away from the catalytic site. Indeed, the connection between domains IV and V consists of two relatively long polypeptide stretches without secondary structure, and the relative positions of domain V in the different GS structures with respect to the rest of the proteins indicates some flexibility in these stretches.

**Domain V**, located next to domain IV, is either built up of both N- and C-terminal segments (GTF180\(-\Delta N\), GTFA\(-\Delta N\)) or of an N-terminal segment alone (DSR-E \(\Delta N_{123}\)-GBD-CD2). Despite this difference, the arrangement of secondary structure elements in domain V is remarkably similar. This can be attributed to the presence of a common structural module (Figure 7a), which is present in multiple copies: one N-terminal module and two C-terminal modules in GTF180\(-\Delta N\) and GTFA\(-\Delta N\), or three N-terminal modules in DSR-E \(\Delta N_{123}\)-GBD-CD2. The structural modules contain two or three consecutive ‘β2/β3 units’ of about 20 residues (Figure 7b); each unit consists of a β-hairpin or a 3-stranded antiparallel β-sheet. Together, the β2/β3 units are arranged in a regularly repeating fashion, resulting in a β-solenoid fold. Structural variation of this fold is obtained by the presence of short additional β-strands or β-hairpins, or by the ‘antiparallel doubling’ of β-hairpins from different modules (in GTF180\(-\Delta N\) and GTFA\(-\Delta N\) but not in DSR-E \(\Delta N_{123}\)-GBD-CD2). The β-solenoid modules are not only structurally superimposable but also display sequence similarity (Figure 7c). Moreover, because each module contains two or three β2/β3 units, they also contain internal repeating sequence motifs. For example, the second β-strand in each unit contains one or more aromatic residues (at least one tyrosine), and the following hairpin loop has a glycine residue. Furthermore, successive β2/β3 units within a module are separated by a basic residue (lysine or arginine).
Figure 7. Domain V. Organization and combined structure/sequence alignment of domain V observed in the structures of GTF180-ΔN, GTFA-ΔN and DSR-E Δ123-GBD-CD2. The β2/β3 motifs (see text) usually contain either a 2-stranded β-hairpin (orange) or a 3-stranded β-sheet (green); extra β-strands are shown in cyan. Two or three motifs form a common β-solenoid structural module. (a) Superposition of structural modules. (b) Schematic organization of domain V and its relative location in the glucansucrases; boxes correspond to motifs. (c) Structural/sequence alignment of domain V with the aligned sequence of GTF-SI domain V below. Secondary structure assignments are based on DSSP (Kabsch & Sander 1983). Open boxes correspond to the YG-repeats (black lines), A-repeats (blue lines, residues) and C-repeats (red lines, residues); YG motifs are shown in bold. Grey boxes correspond to the basic residues between motifs. Residues in grey are not visible in the structure.

Previously, different types of sequence repeats have been identified in the N- and C-terminal GS domains, such as the YG-repeat containing a tyrosine/glycine motif and the A-/B-/C-/D-types of repeats (Bozonnet et al., 2002; Giffard & Jacques 1994; Janeček et al., 2000; Kralj et al., 2004a,b). It appears that in GSs containing these repeats, a common building block is used to build up domain V; variation between GSs is obtained from differences in the number of building blocks, whether they are present in the N-terminal
and/or C-terminal segments, and the way they are connected and packed. Whether the same building blocks are present in GSs with longer N- or C-terminal segments is currently unknown. In this light, the identification of several types of sequence repeats in those segments suggests that the modular structure observed in domain V may be extended. Furthermore, the lateral packing of N- and C-terminal modules observed in the *Lactobacillus* GS structures may be common for GSs with both N- and C-terminal segments outside the catalytic core.

### 2.5. Glucan binding domains

The predicted N- and C-terminal segments of GSs have been designated ‘glucan binding domains’ (GBDs). Based on truncation and deletion studies (Abo *et al.*, 1991; Brison *et al.*, 2012; Kingston *et al.*, 2002; Kralj *et al.*, 2004a,b; Lis *et al.*, 1995; Monchois *et al.*, 1999a,b), the GBDs of GSs are thought to be involved in binding their polymerization products, constituting functional domains that can bind glucans by themselves, but that are not directly involved in catalysis; the location of domain V in the GS structures with respect to the catalytic domain A seems to support this hypothesis. Instead, GBDs may facilitate glucan transfer from/to the active site (Van Hijum *et al.*, 2006). Because GBDs have sequence repeats that show homology to carbohydrate-binding motifs in *Clostridium difficile* toxins (Von Eichel-Streiber *et al.*, 1992), it was proposed early on that these repeating elements may represent the sites for glucan binding. The presence of multiple binding sites may enhance the affinity for glucans. Indeed, mutation of conserved aromatic and basic residues in repeating elements of the GBD of *S. downei* GtfI (Shah *et al.*, 2004) indicated that they are important for binding dextran.

Within the A-/B-/>C-/D-repeats identified in many GSs, a consensus YG-repeat was identified (Giffard & Jacques, 1994), which has homology with the cell wall binding motif found in choline binding proteins, toxins and surface-associated proteins. Available 3D structures of such proteins also display a β-solenoid fold, where tandem cell wall binding repeats form the binding sites for *e.g.* choline (*Streptococcus pneumoniae* autolysin LytA choline-binding domain (Fernández-Tornero *et al.*, 2001)) or a trisaccharide (*C. difficile* toxin A C-terminal domain (Greco *et al.*, 2006)) present in cell wall components like (lipo)teichoic acids. Based on the sequence homology with these proteins, Olvera *et al.* (2007) constructed a model for the dextranucrase DsrP C-terminal segment, which displayed a β-solenoid architecture. The 3D structures of GTF180-ΔN, GTFA-ΔN and DSR-E ΔN123-GBD-CD2 confirm that the YG- and A-/C-repeats present in GSs form domains with the same β-solenoid fold. GSs are extracellular enzymes, but sometimes can also be found in a cell-associated form (Van Geel-Schutten *et al.*, 1999). Thus, the GBDs of GSs may, in addition to their ability to bind glucan products, also be able to anchor to carbohydrate moieties present in cell wall components. Detailed structural information
for carbohydrate binding by β-solenoid domains of GSs has yet to be obtained, as soaking experiments with GTF180-ΔN or DSR-E ΔN₁₂₃-GBD-CD2 were unsuccessful (unpublished results; Brison et al., 2012).

The crystal structures of GSs only reveal part of the GBDs (Figure 4). For both GTF180-ΔN and GTFA-ΔN, the N-terminal ~700-residues are absent from the construct, whereas in DSR-E ΔN₁₂₃-GBD-CD2 the first ~640 of the predicted 839 residues of the GBD are not present. These missing stretches do contain sequence repeats and may extend the β-solenoid fold observed in domain V. The structure determination of a complete GS therefore remains an interesting challenge. Attempts to crystallize full-length L. reuteri glucansucrases GTFA and GTF180 in our laboratory have been successful (T. Pijning, unpublished results), but the diffraction limit of the crystals (7-8 Å resolution at best) was not sufficient to determine their crystal structure. However, low-resolution structural information could be obtained from small-angle X-ray scattering (SAXS) experiments (Chapter 3). These first structural data for a full-length GS (Figure 9c in Paragraph 2.6.) reveal an extension of the elongated crystal structure, resulting in an almost symmetrical boomerang-like molecular shape with a bend halfway. A precise assignment of the location of the domains in this shape was not possible, but it seems plausible that the ~700-residue N-terminal domain(s) extend away from domain V, with which they form one ‘arm’ of the boomerang (~900 residues in total). The other arm would then harbor domains C, A, B and IV (~840 residues).

2.6. Flexibility of domain V

How GBDs of GSs facilitate the transfer of synthesized products from the catalytic site is unknown; it has been speculated that domain V may ‘swing’ between two conformations such as to bring bound intermediate glucan products away from or towards the active site (Ito et al., 2011) but to date no experimental evidence exists for such movement. However, the recent crystal structures do shed some light on this matter.

As stated before, the contacts between domains IV and V in GTF180-ΔN and GTFA-ΔN are relatively small, and domain V takes up somewhat different positions in the various structures with respect to the rest of the GS molecule. Such a ‘hinge movement’ gives rise to shifts of about 20 Å at the ‘top’ of domain V (Figure 8a). In another crystal form of GTF180-ΔN (unpublished data), a 12° rotation of domain V results in an 8 Å shift. In both GTF180-ΔN and GTFA-ΔN, domain V is located such that the overall structure has an elongated shape with a longest axis of about 150 Å. For multi-domain proteins, small domain movements are not uncommon, and may arise from packing effects in crystal lattices. Surprisingly however, in DSR-E ΔN₁₂₃-GBD-CD2, domain V is relocated completely such that it packs against the side of domains IV and B. This domain packing results in a more compact overall protein structure with a longest axis of ~115Å.
Supporting evidence for such pronounced flexibility of the domain IV – domain V hinge comes from a third, orthorhombic crystal form of GTF180-ΔN (Figure 8b; PDB: 4AYG, Chapter 3) which shows the same compact domain arrangement as DSR-E ΔN_{123}-GBD-CD2 and an unchanged β-solenoid structure of domain V itself.

Figure 8. Flexibility of domain V. (a) Superposition of the structures of GTF180-ΔN in triclinic crystal form (PDB: 3KLK, orange), second triclinic crystal form (unpublished data, green) and GTFA-ΔN (PDB: 4AMC, Pijning et al., 2012 and Chapter 4 of this thesis, red); the locations of the hinges are indicated by stars (⋆). (b) Superposition of the structures of GTF180-ΔN in the triclinic crystal form (PDB: 3KLK, orange) and GTF180-ΔN in orthorhombic form (PDB: 4AYG, Chapter 3, magenta).

The available crystal structures thus clearly suggest that intrinsic flexibility exists in different types of GSs, such that they may hinge between an elongated and a compact conformation. Whether this observed flexibility is an essential property for the function of GSs cannot be determined from (static) crystal structures. In contrast, the small-angle X-ray scattering (SAXS) technique is capable of studying flexibility, since it allows determination of molecular shapes in solution - albeit at lower resolution limits than X-ray crystallography. The recently determined solution structure of GTF180-ΔN (Chapter 3) shows that in solution, the enzyme adopts a single elongated conformation. This conformation fits well with the elongated crystal structure (Figure 9a), but not with the compact crystal structure (Figure 9b). Also the solution structure of full-length GTF180 determined by SAXS (Figure 9c) reveals an elongated shape with no indications of extensive flexibility. In the presence of the substrate sucrose, when early products (oligosaccharides) are being synthesized, only minor changes in the molecular shape are observed. In conclusion, although GTF180 and other GSs have an intrinsic flexibility, no large hinging motion occurs in solution, not even when the enzyme actively processes its substrate into short oligosaccharides.
Chapter 1

Figure 9. Solution structures from SAXS experiments (see Chapter 3). (a,b) GTF180-ΔN SAXS molecular shape superimposed with the elongated and compact crystal structures, respectively. (c) Molecular shape of full-length GTF180.

2.7. Catalytic mechanism

Glucansucrases transfer the glucose moiety of sucrose to an acceptor molecule. Based on the type of acceptor substrate used three types of reactions can be distinguished (Figure 10). The dominant reaction is polymerization yielding dextran, mutant, alternan or reuteran; elongation of the α-glucan takes place at the non-reducing end. Especially at low acceptor substrate concentrations (i.e. at the start of a sucrose conversion), however, GSs also utilize water as acceptor substrate, simply hydrolyzing sucrose to glucose and fructose. The third type of reaction is the acceptor reaction, in which glucose is transferred to acceptor molecules other than a growing α-glucan chain.

Figure 10. The three types of reactions catalyzed by glucansucrases. After cleavage of the substrate sucrose and formation of the covalent glucosyl-enzyme intermediate, (1) hydrolysis, (2) coupling to an acceptor, or (3) elongation at the non-reducing end (NR) (α-glucan synthesis) can take place.
For example, the α(1→4)-linked disaccharide maltose is a known acceptor for several GSs, including *L. reuteri* 121 GTFA and *L. reuteri* 180 GTF180 (Koepsell *et al*., 1953, Kralj *et al*., 2004b, Vujičić-Žagar *et al*., 2010). In addition, several GSs catalyze a side reaction (i.e. promiscuous activity) that involves the transfer of the non-reducing end glucose moiety of an α-glucan chain to another α-glucan chain in a so called disproportionation reaction (Binder *et al*., 1983).

![Figure 11. Reaction mechanism of oligosaccharide/polymer formation by GH70 glucansucrases. Asp is the nucleophilic aspartate, Glu is the general acid/base glutamate, and R represents a sugar group. The substrate sucrose, binds in subsites -1 and +1 (Davies *et al*., 1997) and, after formation of a covalent glucosyl-enzyme intermediate, an α(1→4) glycosidic linkage is formed. Stars (*) indicate the hydroxyl groups that are hydrogen-bonded by the transition-state stabilizing aspartate residue (not drawn).](image)

The first 3D structure of a GH70 glucansucrase, *L. reuteri* 180 GTF180-ΔN and its complexes with sucrose and maltose (Vujičić-Žagar *et al*., 2010), confirmed that they use the same set of key amino acid residues for catalysis and the same mechanism (Figure 11) as the related GH13 and GH77 enzymes, although the latter two act on differently linked α-glucose polymers (Barends *et al*., 2007; Devulapalle *et al*., 1997; Vujičić-Žagar *et al*., 2010; Uitdehaag *et al*., 2002). The role of the catalytic residues in the reaction
mechanism has been extensively studied in GH13 enzymes such as Bacillus circulans cyclodextrin glucanotransferase (Uitdehaag et al., 1999), N. polysaccharea amyllosucrase (Skov et al., 2002) and Aspergillus oryzae α-amylase (Brzozowski & Davies 1997). In the GTF180-ΔN D1025N mutant, the observed interactions of the proposed catalytic residues D1025, E1063 and D1136 with the substrate sucrose (Figure 12a) are fully compatible with the proposed double-displacement mechanism. The substrate sucrose is cleaved, resulting in a covalent β-glucosyl-enzyme intermediate (a model is shown in Figure 12b). In the second half-reaction, the glucosyl moiety is then transferred to an acceptor with retention of the α-anomeric configuration (Koshland, 1953; Uitdehaag et al., 1999).

Of the 7 conserved GH13 residues (Uitdehaag et al., 2002), the 3 catalytic residues, as well as 3 other strictly conserved residues in GH70 enzymes (R1023, H1135, and D1504, which is hydrogen bonded to the stacking Y1465) bind and orient the substrate such as to favor formation of the covalent intermediate in the deep pocket forming subsite -1. Additionally, a glutamine residue (Q1509) found in all GH70 GSs replaces the conserved histidine found in GH13 enzymes but performs the same function. The active site of GSs is ‘blocked’ beyond subsite -1 (by Q1140, N1411, D1458); such a pocket-like shape is also observed in the N. polysaccharea amylosucrase. Due to this feature, GSs can transfer only a single glucose moiety per reaction cycle, but not oligosaccharides, in contrast to the α-amylases, which have a longer binding groove.

Figure 12. (a) Structure of GTF180-ΔN mutant D1025N in complex with sucrose (Vujčić-Žagar et al., 2010). The substrate is shown with yellow carbon atoms; residues from domain A are shown with blue carbon atoms, and residues from domain B with green carbon atoms. Hydrogen bonds are shown as dashed lines. (b) Model of the covalent glucosyl-enzyme intermediate in GTF180-ΔN, based on the crystal structure of a N. polysaccharea amylosucrase covalent intermediate (Jensen et al., 2004). The glucosyl moiety is shown with orange carbon atoms; residues from domain A are shown with blue carbon atoms, and residues from domain B with green carbon atoms.
2.8. Product specificity

The various GSs synthesize different products. For example, GTF180-ΔN synthesizes dextran with mainly α(1→6), GTFA-ΔN reuteran with mainly α(1→4), and GTF-SI mutan with mainly α(1→3) linkages. The CD2 domain of DSR-E catalyzes the formation of α(1→2) branches on dextran (and thus is not a glucansucrase sensu stricto) (Bozonnet et al., 2002). Recently, an overview of current insights regarding GS bond specificity has been published by Leemhuis et al. (2013a); so far, some residues and/or regions have been identified that seem to be important in this regard.

The clear differences in GS product linkage type specificities must originate from amino acid or conformation differences in the acceptor regions, i.e. the positive substrate binding subsites according to the nomenclature of Davies et al. (1997) of the various GSs. It is likely that amino acid residues in the acceptor region are responsible for the precise orientation of an acceptor in subsite +1, and thus are determinants of the type of α-glycosidic linkage formed to the C1 atom of the glucosyl-enzyme intermediate (1→6, 1→4, 1→3 or 1→2). Before 3D structural information was available, mutation studies were mainly aimed at residues in conserved regions I-IV (Figure 6, page 24); some of these residues are strictly conserved, while others are only moderately conserved. The crystal structures of GTF180-ΔN and GTF-SI with bound maltose revealed that residues interacting with this acceptor in subsites +1 and +2 (Figure 13a,c) are from regions II, III and IV but also from other non-conserved regions. By structural superposition, the corresponding residues in GTFA-ΔN and DSR-E ΔN123-GBD-CD2 can be identified (Figure 13b,d).

The non-reducing end sugar moiety of maltose in subsite +1 is surrounded by residues from domain B and regions II and III, some of which interact directly with the acceptor. Indeed, mutations in regions II and III in GTFA-ΔN affected the product linkage type distribution or the transferase activity, respectively (Kralj et al., 2005); in DSRE ΔN123-GBD-CD2 the region II mutant F2214N was unable to use dextran as an acceptor (Brison et al., 2012). Residues from domain B have not been targeted for mutation so far.

At subsite +2, residues near the reducing end sugar moiety of maltose are from regions III, IV and helix α4. First, from region III, a tryptophan residue present in almost all GSs (W1065 in GTF180-ΔN, W517 in GTFA-SI) has a stacking interaction with the acceptor. Mutation studies in GTF180-ΔN (Kralj et al., 2009; Petersen, 2011) and GTF-I of S. mutans (Tsumori et al., 1997) showed that the presence of this aromatic stacking platform is essential for GS activity. Interestingly, the CD2 of DSR-E (which does not elongate α-glucans but only branches them) lacks this feature, as it is replaced by A2249-G2250 at the corresponding position. Second, in several mutation studies (Hellmuth et al., 2008; Kang et al., 2011; Kralj et al., 2005, 2006; Moulis et al., 2006; Van Leeuwen et al., 2008c, 2009) the importance of region IV residues following the transition state
stabilizer was shown, mainly affecting the linkage type distribution of the products. Likely the residues of this tripeptide region (see Figure 6, blue box, page 24) determine the orientation of the sugar moiety in subsite +2, as they interact with the \(^{3}\text{OH}\) and \(^{6}\text{OH}\) groups of the sugar (Figure 13a,c). The third group of residues near the sugar moiety bound in subsite +2 involves non-conserved residues from helix \(\alpha 4\) that interact with the \(^{2}\text{OH}\) group. This set of residues has not been targeted in mutation studies so far.

Figure 13. Acceptor substrate binding subsites +1 and +2 in glucansucrases. Shown are (a) GTF180-ΔN with maltose (Vujičić-Žagar et al., 2010), (b) GTFA-ΔN (Pijning et al., 2012, Chapter 4), (c) GTF-SI with maltose (Ito et al., 2011) and (d) DSR-E \(\Delta N_{123}\)-GBD-CD2 (Brison et al., 2012). Residues from domain B are shown with green carbon atoms, region II with magenta carbon atoms, region III with cyan carbon atoms, region IV with dark blue carbon atoms, and helix \(\alpha 4\) with orange carbon atoms.
Taken together, the general aromatic stacking platform provides a rather non-specific interaction with acceptor sugar units, while specificity is likely determined by the less conserved surrounding residues from domain B, regions II and III (subsite +1), as well as those from helix α4 and from regions III and IV (subsite +2). How exactly the differences in surrounding residues determine acceptor orientation in different GSs remains to be determined, since maltose is not an α(1→6) linked dextran-type acceptor, and structures of complexes with acceptors resembling (intermediate) GS products are not available yet. Also, many GSs catalyze the formation of two types of glycosidic linkages, requiring different acceptor binding modes. The maltose complexes reveal how a glucose moiety in subsite +1 is oriented to form an α(1→6) glycosidic bond, but in order to form α(1→3) glycosidic linkages or branches in the product, different binding modes must exist. Docking experiments with GTF180-ΔN have shown that such binding modes are feasible (Vujičić-Žagar et al., 2010). The α-glucan produced by GTF180-ΔN contains consecutive α(1→6) linkages, but consecutive α(1→3) glycosidic linkages are never observed (Van Leeuwen et al., 2008c). In case of GTFA, initially formed gluco-oligosaccharide products contains alternating α(1→6) and α(1→4) linkages (Dobruchowska et al., 2013). These results indicate that, in addition to the residues forming acceptor subsites, also the type of linkage between the sugar units in subsites +1 and +2 determines which linkage is formed in the next transglycosylation step. Recently, Leemhuis et al. (2013a) published an overview of current insights regarding GS bond specificity and acceptor binding residues.

At present, it is not clear whether GH70 GSs use a processive or non-processive mechanism. If, in a processive mechanism, intermediate products stay bound to the enzyme, acceptor binding sites beyond +1 and +2 may ‘come into play’ during the course of elongation, when longer oligosaccharides are formed and also function as acceptor molecules in the next elongation or branching step. In this light, it is relevant to compare the surfaces of the enzymes and the structural elements that shape binding grooves near the active site (Figure 14). The surfaces of GTF180-ΔN and GTFA-ΔN, performing mainly α(1→6) and α(1→4) elongation, respectively, both show a rather wide binding cleft (Figure 14a,b). In GTF-SI, which mainly synthesizes α(1→3) linked glucans, one side of the binding cleft is less open due to the presence of a longer loop (region ‘B’, Figure 14c). In DSR-E ΔN123-GBD-CD2 this side of the binding cleft is even more ‘blocked’ by a unique long β-hairpin subdomain and a smaller β-hairpin (region ‘F’ and ‘D’, respectively, Figure 14d). Notably, the CD2 of DSR-E only has α(1→2) branching activity, probably requiring only one acceptor binding mode but still requiring the binding of long dextran-type α-glucans. Thus, the different product specificities of these GSs are reflected in their different binding cleft shapes.

In conclusion, both the overall shape of the binding clefts, as well as key interactions provided by individual amino acid residues, are likely of key importance for GS protein
engineering studies aimed at manipulating their product specificity. Attempts to achieve such goals either by directed evolution or rational mutagenesis have been undertaken (Emond et al., 2007; Irague et al., 2012; Kelly et al., 2009).

**Figure 14.** Comparison of six regions (A, red; B, yellow; C, green; D, cyan; E, blue; F, magenta) surrounding the active site cleft in glucansucrases. (a) GTF180-ΔN with maltose (yellow sticks) (Vujičić-Žagar et al., 2010). (b) GTFA-ΔN (Pijning et al., 2012; Chapter 4). (c) GTF-SI with maltose (yellow sticks) (Ito et al., 2011). (d) DSR-EΔN123-GBD-CD2 (Brison et al., 2012). The covalent glucosyl-enzyme intermediate models are shown in stick representation with white carbon atoms.

### 2.9. Glucansucrases GTF180 and GTFA

Screening of lactic acid bacteria (LAB) revealed several strains able to produce water-soluble α-glucans (and β-fructans) from sucrose (Van Geel-Schutten et al., 1999). The first characterized enzyme responsible for α-glucan production in LAB was the glucansucrase GTFA from *L. reuteri* 121 (Kralj et al., 2002, 2004b). Later, other GSs from different LAB strains were identified and characterized, such as GTF180 from *L. reuteri* 180 (Kralj et al., 2004a). GTFA and GTF180 were studied extensively, using site-directed mutagenesis, hybrid mutagenesis and structural studies on the enzymes and their products. Both enzymes are optimally active around pH 4.5 and 50 °C, and require Ca²⁺ for activity.
Glucansucrase GTFA from *L. reuteri* 121 is a 1781-residue enzyme synthesizing a branched α-glucan with α(1→6) and α(1→4) linkages (Table 2). Deletion of the 742-residue N-terminal domain hardly affected the product, although initial transferase activity decreased. Such a construct, GTFA-ΔN, was the first crystallized GH70 GS but due to the limited diffraction properties of the crystals, its structure could only be determined later (Pijning *et al*., 2012, Chapter 4). Meanwhile, Kralj *et al.* (2005, 2006) showed that linkage specificity could be manipulated by site-directed mutagenesis, e.g. from mainly reuteran towards dextran; especially a short region (residues 1134-1136) in motif IV (Figure 6, page 14) seemed to be crucial in this respect. A composite structure of the ‘final’ α-glucan synthesized by GTFA was determined by Van Leeuwen *et al.* (2008c) (Figure 15a). Recently, also initial products formed by GTFA upon incubation with sucrose (or malto-oligosaccharides) have been analyzed (Dobruchowska *et al*., 2013). Interestingly, elongation of sucrose with glucose units seems to take place via alternating α(1→4) and α(1→6) linkages, yielding linear products up to at least DP12. Moreover, simultaneously with oligosaccharides, polymers are being synthesized from the beginning.

**Table 2.** Composition and size of the products of GTFA and GTF180, either full-length or truncated (ΔN); data apply to recombinant enzymes (taken from Kralj *et al*., 2004a,b).

<table>
<thead>
<tr>
<th>Linkage type occurrence (%)</th>
<th>Enzyme</th>
<th>GTFA</th>
<th>GTFA-ΔN</th>
<th>GTF180</th>
<th>GTF180-ΔN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal</td>
<td></td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>α(1→4)</td>
<td></td>
<td>49</td>
<td>46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α(1→6)</td>
<td></td>
<td>26</td>
<td>34</td>
<td>52</td>
<td>54</td>
</tr>
<tr>
<td>α(1→4,6)</td>
<td></td>
<td>15</td>
<td>12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α(1→3)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td>α(1→3,6)</td>
<td></td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Product MW (MDa)</td>
<td></td>
<td>45</td>
<td>50</td>
<td>36</td>
<td>n.d.</td>
</tr>
</tbody>
</table>

Glucansucrase GTF180 from *L. reuteri* 180 has 78% sequence identity to GTFA (Kralj *et al*., 2004b). Its product is a branched glucan with α(1→6) and α(1→3) linkages (Table 2). Like for GTFA, a construct lacking the N-terminal domain (741 residues, GTF180-ΔN) was fully active. A GTF180-ΔN construct could be crystallized readily (Pijning *et al*., 2008, Chapter 2), leading to the first GH70 GS crystal structure and the discovery of the unusual domain organization of these enzymes. Mutation studies with GTF180 targeting
the same region IV residues as in GTFA revealed that also the linkage type distribution in the α-glucan produced by wild-type GTF180 (Van Leeuwen et al., 2008b) could be altered (Van Leeuwen et al., 2009). The structure of GTF180-ΔN, also in complex with the acceptor maltose, showed that these residues are near acceptor binding subsite +2. Other residues targeted in mutation studies included a conserved tryptophan (W1065) (Kralj et al., 2009; Petersen, 2011). Analysis of initial reaction products is currently in progress; in addition, truncation studies are performed where domains IV and/or V are deleted to study their role in the reaction and determining product specificity (X. Meng, personal communication).

![Figure 15. Composite structures of the α-glucans synthesized by GTFA (a) and by GTF180 (b) (Van Leeuwen et al., 2008b,c). The α-D-GlcP represents one glucosyl unit.](image)

3. FRUCTANSUCRASES

3.1. Classification and primary structure of bacterial fructansucrases

To synthesize β-fructans from sucrose, plants use a two-step route; bacteria on the other hand use a single fructansucrase (FS) enzyme to do so. Corresponding to the type of β-fructan made, they are called levansucrases (making β(2→6)-linked levan) and inulosucrases (making β(2→1)-linked inulin). Levansucrases are found in a variety of both Gram-positive and Gram-negative bacteria. In contrast, inulosucrases are found exclusively in Gram-positive species; they have been identified in all LAB, including *Streptococcus mutans*, *Leuconostoc citreum* CW28, *Lactobacillus reuteri* 121, *Lactobacillus reuteri* TMW 1.106, *Lactobacillus johnsonii* NCC 533 and *Lactobacillus gasseri* and in the non-LAB *Bacillus sp* (Anwar, 2010b). While most of the research has been directed towards levansucrases, from species such as *Bacillus*, *Zymomonas*, *Lactobacillus*, *Streptococcus* and *Gluconacetobacter* (Van Hijum et al., 2006), inulosucrases have received more attention only in recent years.
Although fructansucrases share their substrate specificity with glucansucrases, the two enzyme classes do not at all share sequence similarity. Bacterial FSs are classified in glycoside hydrolase family 68 (GH68) (http://www.cazy.org; Cantarel et al., 2009). They are related to GH32 enzymes, which are found in plant and fungal species, and which catalyze reactions to synthesize or degrade β-fructans (e.g. invertases, inulinases). Together, the GH32 and GH68 enzymes constitute clan GH-J, sharing similar structural features and catalytic machineries.

A phylogenetic analysis of bacterial FSs revealed a clear divergence between the enzymes from Gram-negative and Gram-positive species (Anwar et al., 2010b). On the other hand, among the Gram-positive bacteria, the distinguishing differences between levan- and inulosucrases are less clear. For example, in LAB their sequence identity is >60% (Anwar et al., 2008). The inulosucrases do form a small group, indicating that there must be subtle features that distinguish them from levansucrases. Product linkage type specificity (either β(2→6) or β(2→1)) may thus result from small differences between the two types of FSs, providing opportunities to engineer these enzymes in such a way that product specificity can be changed.

Most microbial levansucrases are monomeric and have a molecular mass between ~45-75 kDa (Van Hijum et al., 2006), but values up to 220 kDa have been reported (Velázquez-Hernández et al., 2009). Inulosucrases in general are larger than levansucrases. Based on their sequence, it was proposed that bacterial FSs contain a central catalytic domain of ~500 residues; this is the most conserved domain and contains 11 conserved sequence motifs. The catalytic domain would be flanked by smaller N- and C-terminal domains. The N-terminal domain, displaying more sequence variability, is preceded by a signal peptide (~30 residues) related to export of the enzyme across the cell membrane, into the extracellular space. In some cases, the N-terminal domain contains sequence repeats (Tieking et al., 2005). The C-terminal domain has a variable region, and sometimes contains an LPXTG motif, e.g. in L. reuteri 121 FSs (Van Hijum et al., 2002, 2004) and in L. johnsonii NCC 533 inulosucrase (Anwar et al., 2008); in these cases the FS is anchored to the bacterial cell membrane.

The exact function of the N- and C-terminal domains is not clear. In the levansucrase SacB from Bacillus subtilis extension of the C-terminal domain was found to affect product branching (Chambert et al., 1992). In other cases, chimeric fructansucrases carrying glucansucrase-like C-terminal domains showed a modulated catalytic efficiency and thermal stability (Morales-Arrieta et al., 2006; Olivares-Illana et al., 2003).

3.2. Structural aspects of GH68 fructansucrases

The first 3D structural data for bacterial FSs became available about a decade ago, for levansucrases BsSacB from the Gram-positive bacterium B. subtilis (Meng & Fütterer,
2003) and GdLsdA from the Gram-negative Gluconacetobacter diazotrophicus (Martinez-Fleites et al., 2005). The structure of the inulosucrase from L. johnsonii NCC 533 (InuJ) (Anwar et al., 2008; Pijning et al., 2010 (Chapter 5)) is the first structure of an inulosucrase. Its catalytic domain is flanked by an N- and a C-terminal domain, which were both truncated (Figure 16).

![Figure 16. Schematic domain organization of the inulosucrase InuJ from L. johnsonii NCC 533. The catalytic domain is shown in yellow with the 5-bladed β-propeller fold and the central cavity containing the active site. The N- and C-terminal domains (both truncated) are shown in cyan and red, respectively.](image-url)

### 3.2.1. Catalytic domain

The catalytic domain of GH68 inulo- and levansucrases (and related GH32 enzymes) is globular in shape and contains a rare 5-bladed β-propeller fold. Each blade consists of 4 antiparallel β-strands (A-D), and together they surround a deep central cavity. This funnel-shaped cavity is lined with conserved residues from the A-strands and several loops. At the bottom of the funnel, a deep pocket exists which is lined with negatively charged amino acid residues. The three invariant GH68 catalytic residues (see Paragraph 3.3.) are located near the bottom of this pocket, which is highly conserved between levan- and inulosucrases.

Most of the sequence variation between different FSs is found in the elements that connect the β-strands in the blades of the β-propeller fold, especially in the loops that connect blades B and C. For example, in inulosucrase InuJ, this connection contains an α-helix at the rim of the active site funnel that is absent in levansucrases BsSacB and GdLsdA.

The catalytic domain of FSs from Gram-positive bacteria harbors a conserved Ca$^{2+}$ binding site not far from the active site pocket; the Ca$^{2+}$ ion is coordinated by residues from loops connecting β-strands of blades 3 and 4. These residues participate in a hydrogen bonding network that links to the catalytic site; therefore the Ca$^{2+}$ site not only maintains local structure but also is essential for FS activity, as was already found before structures were known (Ozimek et al., 2005; Van Hijum et al., 2006).
3.2.2. N- and C-terminal domains

The N- and C-terminal domains of GH68 FS structures are packed to the sides of the β-propeller domain (Figure 16). While the N-terminal stretch contains several α-helices, the C-terminal stretch contains hardly any secondary structure (except for a small 2-stranded β-sheet in InuJ). The N- and C-terminal stretches are located far from the active site, and thus not likely to be involved directly in the FS transferase reaction. However, these domains could in principle provide remote binding sites for (growing) levan- or inulin-type FOS and polymers.

3.3. Mechanism of bacterial fructansucrases

Based on the 3D structures and various mutagenesis studies with bacterial FSs (Chambert & Petit-Glatron, 1991; Meng & Fütterer, 2003, 2008; Ortiz-Soto et al., 2008; Ozimek et al., 2006a; Yanase et al., 2002), a two-step double displacement reaction mechanism has been proposed, which involves formation of a fructosyl-enzyme intermediate (Figure 17). This ping-pong type reaction mechanism has been demonstrated for FSs from *B. subtilis, G. diazotrophicus* and *Streptococcus salivarius* (Chambert et al., 1974; Hernández et al., 1995; Song & Jacques, 1999). It involves a single active site, located in a deep pocket, such that the donor substrate (sucrose) is bound in subsites -1 and +1 (Davies et al., 1997). The first GH68 inulosucrase structure (Pijning et al., 2011, Chapter 5) revealed a very similar active site architecture and catalytic machinery, thus confirming that GH68 levan- and inulosucrases have a conserved substrate specificity (for sucrose) and employ the same mechanism.

The FS mechanism (Figure 17) involves three conserved amino acid residues: two aspartate residues and a glutamic acid residue. In the first step, cleavage of the donor substrate sucrose is promoted by protonation of the glycosidic bond by the glutamic acid residue, resulting in the formation of an oxocarbenium-ion type transition state. An aspartate residue contributes to stabilization of this transition state, via hydrogen bonds to the C3 and C4 hydroxyl groups. The partially positively charged (δ+) anomeric carbon atom of the transition state undergoes a nucleophilic attack by the second aspartate residue. Next, the glycosidic bond is cleaved, whereby the fructosyl moiety is covalently coupled to the nucleophilic aspartate, and glucose is released. In the second half of the reaction, the covalently linked fructosyl moiety is transferred either to water (hydrolysis) or to an incoming acceptor sugar (transfructosylation); this occurs again via an oxocarbenium-ion type transition state. The product has a retained anomeric configuration (at carbon C1).
Figure 17. Reaction mechanism of oligosaccharide/polymer formation by GH68 FSs; Asp is the nucleophilic residue, Glu is the general acid/base residue, and R represents a sugar group (if the acceptor is sucrose, then R is a glucosyl moiety). In this example, a $\beta(2\rightarrow1)$ linkage is formed giving an inulin-type product; alternatively, the acceptor can be attached to the 6-OH resulting in a $\beta(2\rightarrow6)$ linkage in the case of levan synthesis. Stars ($\star$) indicate the hydroxyl groups that are hydrogen-bonded by the second aspartate residue (not drawn); the anomeric carbon is indicated in the substrate and product with a black circle ($\bullet$).

Figure 18. The three types of reactions catalyzed by FSs. After cleavage of the substrate sucrose and formation of the covalent fructosyl-enzyme intermediate, (1) hydrolysis, (2) coupling to an acceptor, or (3) $\beta$-fructan elongation by coupling to another sucrose molecule, may occur.

The acceptor sugar can be either sucrose - which leads to the synthesis of $\beta$-fructan oligosaccharides or polymers - or another type of sugar (Figure 18). For example, several levansucrases have been shown to be able to use maltose, maltotriose, raffinose,
General introduction

cellobiose etc. Moreover, it has been shown that sucrose analogues may function as acceptor substrates (Kralj et al., 2008a); this then provides the possibility to synthesize novel FOS. Such carbohydrates may be applied as prebiotics in food and health products, or they may be used as building blocks for the synthesis of glycosylated drugs or pharmaceutical products (Beine et al., 2008; Homann & Seibel, 2009a,b; Kralj et al., 2008a).

Based on studies with the levan- and inulosucrase from L. reuteri 121, Ozimek et al. (2006b) proposed a model for β-fructan synthesis. First, the substrate sucrose binds at subsites -1 and +1 (nomenclature according to Davies et al., 1997). Its glycosidic bond is cleaved, resulting in a covalent fructosyl-enzyme intermediate. A second sucrose molecule then enters the active site, occupies subsites +1 and +2, and forms a new glycosidic bond with the fructosyl-enzyme. Further elongation of the first trisaccharide product (kestose) takes place by adding new fructosyl units at the reducing end, each derived from sucrose. Since the active site is a ‘dead end’ funnel, after each elongation step, the intermediate product has to ‘shift’ to higher subsites in order to make room for an incoming donor substrate sucrose molecule.

3.4. Substrate- and product specificity

Fructansucrases synthesize a wide variety of products, differing in size (DP), linkage type (β(2→1) or β(2→6)), branching amount etc. To understand the differences in substrate- and product specificity of FSs, knowledge of enzyme structures - also in complex with substrates, acceptors and products - is of key importance. Several structures of GH68 and of the related GH32 enzymes have been determined, the first ones being structures of the bacterial levansucrase SacB from B. subtilis (BsSacB (Meng & Fütterer, 2003)). Structures of GH32 yeast and fungal enzymes have also been reported (Álvaro-Benito et al., 2010a, b; Álvaro-Benito et al., 2012; Chuanhkayan et al., 2010; Nagem et al., 2004; Pouyez et al., 2012; Sainz-Polo et al., 2013). The groups of Seibel (Braunschweig / Würzburg) (Homann et al., 2007; Seibel et al., 2006; Strube et al., 2011) and Dijkhuizen / Dijkstra (Groningen) (a.o. Anwar et al., 2012; Kralj et al., 2008a,b; Ozimek et al., 2004, Ozimek et al., 2006a; Pijning et al., 2011; Van Hijum et al., 2004) focus on GH68 bacterial levan- and inulosucrases. Together, the 3D structures have revealed some of the principles of FS specificity.

In GH68 FSs, substrate specificity is shared by levan- and inulosucrases. In both enzyme types, the binding mode of the donor substrate sucrose is the same, especially the binding mode of the fructosyl moiety in subsite -1 (Chapter 5: Figure 6b). A very conserved set of residues orients the disaccharide in a specific way, resulting in an ideal binding mode for cleavage of the glycosidic bond. This first step thus proceeds very similar in both enzyme types.
On the other hand, product size specificity is much more variable; it is determined when the enzyme-linked fructosyl unit is transferred either to water or to an incoming acceptor substrate. Regarding size specificity, it is understandable that sucrose concentration, the architecture of subsites -1 and +1, and acceptor affinity at higher subsites are important. First, at higher sucrose concentrations, oligosaccharide/polymer formation is favored over hydrolysis; the concentration of the donor substrate thus affects the transfructosylation/hydrolysis (T/H) ratio and the size of the final products. Second, the architecture of the active site and its accessibility for water also affect the T/H ratio and product size, as was demonstrated by point mutations in conserved motifs near the sucrose binding site. For example, in *L. reuteri* 121 inulosucrase, mutations in subsite -1 resulted in an enzyme synthesizing less oligosaccharides and more polymer (Ozimek *et al.*, 2006b); in a later study, mutations near the sucrose-binding site affected the T/H ratio (Anwar *et al.*, 2012). In *BsSacB*, mutations near the sucrose-binding site resulted in an enzyme no longer able to produce levan polymer but only oligosaccharides (Ortiz-Soto *et al.*, 2008). Third, product size specificity depends on the affinity of the enzyme for intermediate products near (but not in) the active site, while another donor substrate (sucrose) is being cleaved. Several studies confirm the importance of this ‘remote affinity’ concept. For example, mutating a semi-conserved asparagine residue in the 3B-3C loop (N242 in *BsSacB*, N252 in *BmSacB*) affected product size (Beine *et al.*, 2008, Homann *et al.*, 2007, Homann & Seibel, 2009b); mutation of Y247 in *BmSacB* affected oligosaccharide chain length (Strube *et al.*, 2011). Also an extensive mutational study in *L. reuteri* 121 inulosucrase, focusing on residues that are conserved in inulosucrases but not in levansucrases, revealed that these residues play a role in determining product size (Anwar *et al.*, 2012). Interestingly, the complex of InuJ with sucrose (Pijning *et al.*, 2011; Chapter 5) revealed a secondary sugar-binding site near residues from the 1B-1C loop; mutation of these residues decreased the overall transfructosylation activity of the enzyme. Together, residues further away from the active site seem to contribute to acceptor binding and product size specificity.

Enzymes able to maintain binding and elongating the same intermediate sugar chain until a long product (polymer) has been synthesized are termed processive. In contrast, in non-processive enzymes, having less affinity for intermediate products, the products diffuse out of the enzyme active site more easily, and only shorter oligosaccharides are formed. Some enzymes (*e.g.* InuJ) produce both oligosaccharides and polymer, making the distinction between processive and non-processive enzymes less clear. Together, the interplay between structural factors and reaction conditions determine the range of product size in any FS. Besides product size specificity, linkage type specificity must be governed by the way acceptor substrates bind in the transfructosylation step. In order to form either a β(2→1) or β(2→6) linkage, the terminal fructosyl unit of the acceptor must be oriented in quite different ways in subsite +1. Although several studies using GH68
and GH32 enzymes have addressed linkage type preference, it is still poorly understood. The general emerging picture seems to be that in addition to critical residues near subsite +1, residues from different loops, especially those connecting β-strands B and C in different blades also contribute to acceptor binding at remote sites. These loops form the rim of the central funnel leading to the active site (Chapter 5: Figure 14), and are highly variable in sequence and length. Structural studies of both GH68 and GH32 enzymes have provided the following observations.

**Figure 19.** Sequence alignment of selected conserved loop motifs (I, II, V-VIII) in GH68 bacterial FSs. Lj_Inu, L. johnsonii NCC 533 inulosucrase; Lr_Inu, L. reuteri 121 inulosucrase; Lr_Lev, L. reuteri 121 levansucrase, Bs_Lev, B. subtilis levansucrase; Bm_Lev, B. megaterium levansucrase; Gd_Lev, G. diazotrophicus levansucrase (the only Gram-negative species for which a structure is available). Black triangles indicate the catalytic residues; boxed numbers indicate subsites. Colored boxes represent loops between β-strands B and C in the blades. Figure prepared with ESPript (Gouet et al., 1999); the alignment of the 5B-5C loop was slightly modified based on structural superposition.

For GH68, the complex between inulosucrase InuJ and 1-kestose (βFru(2-1)βFru(2-1)αGlu) (Chapter 5: Figure 7) reveals how the first inulin-type transfructosylation product (with a β(2→1) linkage) interacts in an inulosucrase. Structural information of a GH68 levansucrase complexed with a levan-type oligosaccharide, which would allow a direct comparison of linkage types, is lacking; the trisaccharide raffinose (βFru(2-1)αGlc(6-1)αGal) bound in BsSacB (Meng & Fütterer, 2008), although it is an acceptor, is neither levan- nor an inulin-type. Nevertheless, it is interesting to compare the residues in the two types of enzymes, especially those near subsite +1 (and +2). Interestingly, these residues are almost completely conserved in GH68 FSs (Figure 19, 20). For example, interacting with the sugar moiety at subsite +1, an almost strictly conserved arginine in the 4B-4C loop (e.g. R542 in InuJ, R360 in BsSacB) has been the subject of mutation in bacterial FSs, affecting mainly product size but not linkage type (Homann et al., 2007; Ortiz-Soto et al., 2008).
Figure 20. Acceptor trisaccharide complexes of GH68 and GH32 FSs. InuJ, L. johnsonii NCC 533 inulosucrase (PDB: 2YFT); BsSacB, B. subtilis levansucrase (PDB: 3BYN); AjFT, Aspergillus japonicus fructosyltransferase (PDB: 3LDR); 6-SST/6-SFT: Pachysandra terminalis fructosyltransferase (PDB: 3UGG). The subsites are indicated; nearby residues are shown in stick representation (catalytic residues with orange carbon atoms). Residues from different loops are colored red (1B-1C), green (3B-3C), cyan (4B-4C) and purple (5B-5C), corresponding to the colors in Figure 19.

A second arginine (R545 in InuJ) is less conserved, and mutation of the equivalent K373 in BmSacB affected the DP of the products (Strube et al., 2011). Other mutation studies, also involving residues more remote from the active site, only reported effects on the product size but not on linkage type. Thus, so far it has not been possible to pinpoint residues in GH68 FSs that determine whether the synthesized β-fructans have inulin- or
levan-type linkages. Interestingly, many levan-sucreases from Gram-negative bacteria, while synthesizing a levan-type polymer, also synthesize short inulin-type oligosaccharides (1-kestose, 1,1-nystose and 1,1,1-kestopentaose) (Banguela & Hernández, 2006; Euzenat et al., 2006; Korakli et al., 2003; Támbara et al., 1999). Conceivably, being poor acceptors, such oligosaccharides accumulate without being processed further.

In addition to the GH68 structures, structural information also comes from a few GH32 enzymes. For example, structures of the fungal β(2→1)-fructosyltransferase from Aspergillus japonicus (AjFT) (Chuankhayan et al., 2010) reveal that inulin-type oligosaccharides are bound in a similar orientation as in Inul (Figure 20c). However, the interacting residues near subsites +2 and +3 have no counterpart in GH68 enzymes. A second example is the plant fructosyltransferase 6-SST/6-SFT from Pachysandra terminalis, which synthesizes linear and branched levan-type β-fructans (Lammens et al., 2012). Its complex with 6-kestose in subsites +1 to +3 represents a genuine acceptor binding mode (Figure 20d). Although a direct comparison with an inulin-type acceptor binding mode in the same subsites is not yet at hand, another complex of 6-SST/6-SFT, with 1-kestose bound to subsites -1 to +2 (Lammens et al., 2012), reveals two important structural features for specificity which may also apply to other GH32 FSs, and a comparison with GH68 FSs reveals some interesting points. First, the authors propose that a set of hydrophobic residues from the 1B-1C loop (a.o. residue W57, stacking with the sugar moiety at subsite +2) and from motif II contribute to the predominant β(2→6) linkage specificity of this enzyme. An equivalent aromatic stacking residue is absent in GH68 FSs; usually an arginine or lysine from the 5B-5C loop is at the equivalent position, but there is no clear distinction between levan- and inulosucrases (Figure 19). The second feature proposed to contribute to specificity is an aspartate/glutamine (D244/Q247) couple in the 4B-4C loop near subsite +1; the same was proposed for β-fructofuranosidase from Schwanniomyces occidentalis (Álvaro-Benito et al., 2012). In most GH68 enzymes, this D/Q couple is replaced by an E/R couple from different structural elements; for example E522/R542 near subsite +1 in Inul may have a similar role (Figure 20a). Notably, in several levan-sucreases, mutation of the residue corresponding to R542 in Inul affected the ratio between oligosaccharide and polymer formation. This residue is thus one of the key residues in both GH32 and GH68 FSs.

Despite the insights obtained on GH68 and GH32 FSs, a clear explanation of the differences in linkage type specificity between inulo- and levan-sucreases is still needed. Apart from enzyme architectures, it may be that the structural characteristics of the (intermediate) products also play a role. The different types of flexibility and helical preference between β(2→1) and β(2→6) linked oligosaccharides (Vijn & Smeekens, 1999) likely affect their binding mode and orientation, depending on the shape and surface properties of the enzyme.
3.5. Inulosucrase InuJ from *Lactobacillus johnsonii* NCC 533

*Lactobacillus johnsonii* NCC 533 is an intestinal bacterial strain to which probiotic properties have been ascribed; the food company Nestlé uses it in a yogurt-like dairy product. Anwar *et al.* (2008) characterized the single fructansucrase gene of this organism, which had been annotated as a levansucrase precursor, and showed that it actually encodes an inulosucrase enzyme (InuJ). Sequence analysis revealed that InuJ, a 797-residue protein, contains a predicted N-terminal domain (172 residues), a catalytic core domain (453 residues) and a C-terminal region containing a cell wall anchoring segment (41 residues) and a variable domain (89 residues). Sequence identities to the well-characterized inulosucrases from *L. reuteri* 121 and *L. gasseri* are 60% and 82%, respectively.

A fully active, truncated form of InuJ, InuJΔ144-709His (containing residues 144-709) was biochemically characterized, revealing that it synthesizes a range of inulin-type oligosaccharides (DP ~2-15) as well as linear inulin polymer with a molecular mass of 40 MDa (DP ~250,000). The optimal activity of the recombinant enzyme is at pH 7.0, and at 55 °C; furthermore, the enzyme requires Ca²⁺ ions for activity. Crystallization and structure determination of InuJΔ144-709His provided the first 3D structural information for a GH68 inulosucrase (Pijning *et al.*, 2011; Chapter 5). Later, the structure of InuJΔ144-709His was used to guide and interpret the results of a mutation study in the inulosucrase from *L. reuteri* 121 (Anwar *et al.*, 2012).

4. SCOPE OF THIS THESIS

The research described in this thesis focuses on structural aspects of glucansucrase (GS) and fructansucrase (FS) enzymes from *Lactobacillus* species. These enzymes are able to synthesize molecules that are sometimes 100x bigger (in terms of molecular mass) than the enzymes themselves. Moreover, given the observed polymer structures, and the observation that each GS or FS synthesizes a specific α-glucan or β-fructan type, these enzymes must be specific in the way they put together the building blocks into molecules with defined glycosidic bonds, branching patterns, sizes, etc. Since sucrases elongate their products by attaching one glucosyl or fructosyl unit per reaction cycle, a vast number of cycles is needed to obtain the final EPS product. Research questions regarding GSs and FSs are therefore:

1) What is their reaction mechanism?
2) Are all the reaction steps performed by a single catalytic site?
3) What defines the specificity of elongation?
4) Is elongation processive or non-processive?
These questions can be summarized in one general question: what are the structure-function relationships of glucan- and fructansucrases? To answer this, the structures of these enzymes must be determined at the atomic level. Until recently, structural information was not available for GSs, and in the case of FSs it was limited to levan-synthesizing sucrases. In this thesis, the crystal structures of two LAB GSs and one inulin-synthesizing fructansucrase are described. These structures have allowed us to accurately answer the first two questions, while important insights for product specificity (the third question) could be obtained. In addition, other research groups have studied glucan- and fructansucrases by sequence analysis, mutations at the amino acid level, hybrid mutants, (partial) domain truncation, and analysis of the products. By combining the results from such studies with the structural information, important progress has been made in recent years; still, more research needs to be done to resolve the way these intriguing enzymes work.

**Chapter 1** (this chapter) provides a general introduction to glucansucrases and fructansucrases and their products, and reviews the current knowledge about their structures, mechanisms and specificities.

**Chapter 2** reports on the glucansucrase GTF180 from *L. reuteri* 180, an enzyme that synthesizes α-glucans with α(1→6) and α(1→3) linkages. An N-terminally truncated construct of GTF180 (GTF180-ΔN) was crystallized in three different crystal forms; their crystallographic characterization, along with biochemical characterization of the enzyme are described. The N-terminally truncated enzyme is fully active and shows similar pH and temperature optima as the full length enzyme. All three crystal forms diffract to high resolution. GTF180-ΔN is the first crystallized glycosyl hydrolase family 70 (GH70) glucansucrase, and the first for which the structure was reported (Vujičić-Žagar et al., 2010).

In **Chapter 3**, we describe a second crystal structure of glucansucrase GTF180-ΔN; due to a hinged-back domain V, this structure is much more compact than the elongated structure reported by Vujičić-Žagar et al. (2010). Despite the intrinsic flexibility deduced from the crystal structures, the compact conformation of GTF180-ΔN is not observed in solution; only the elongated conformation appears to be present in solution. In addition, the first structural data for a full-length glucansucrase, GTF180, is presented. The molecular shape reveals the location of the large N-terminal domain that was absent in the truncated GTF180-ΔN.
Chapter 4 describes the crystal structure of the α(1→6) and α(1→4) specific glucansucrase GTFA-ΔN from *L. reuteri* 121, which is highly homologous to GTF180-ΔN. Comparison between the two enzymes reveals the importance of residues in a conserved sequence motif for product linkage type specificity.

Chapter 5 presents the first published crystal structure of a glycosyl hydrolase family 68 (GH68) inulosucrase, InuJ from *L. johnsonii* NCC 533. The complex of InuJ with the substrate sucrose confirms that levan- and inulosucrases share their substrate specificity. The binding mode of the first transfructosylation product 1-kestose reveals for the first time how inulin-type FOS bind in GH68 fructansucrases, and how an inulin-type linkage can be formed. From the structures, we propose that the product linkage type specificity of fructansucrases is determined by residues distant from the active site.

Chapter 6 summarizes the main conclusions of the research described in this thesis. Together with the findings from biochemical and functional experiments, the lessons learned from the structures of glucansucrases and fructansucrases provide directions for future research on these intriguing enzymes.